2013 13th International Conference on Control, Automation and Systems (ICCAS 2013)
Oct. 20-23, 2013 in Kimdaejung Convention Center, Gwangju, Korea

Continuous Critic Learning for Robot Control in Physical Human-Robot Interaction

Chen Wang[1,2], Yanan Li[1,3], Shuzhi Sam Ge[1,2*], Keng Peng Tee[4] and Tong Heng Lee[2]

[1]Social Robotics Laboratory, Interactive Digital Media Institute, National University of Singapore, Singapore 119613

[2]Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119077

[3] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 119613

[4] Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632

(Tel: +65-6516 6821, Fax: +65-6779 1103, E-mail: samge@nus.edu.sg) * Corresponding author

Abstract: In this paper, optimal impedance adaptation is investigated for interaction control in constrained motion. The external environment is modeled as a linear system with parameter matrices completely unknown and continuous critic learning is adopted for interaction control. The desired impedance is obtained which leads to an optimal realization of the trajectory tracking and force regulation. As no particular system information is required in the whole process, the proposed interaction control provides a feasible solution to a large number of applications. The validity of the proposed method is verified through simulation studies.

Keywords: robot-environment interaction; continuous critic learning; impedance adaptation

1. INTRODUCTION

In the predictable future, robots are expected to become a part of the society and collaborate with humans. To realize safe and efficacious human robot interaction in a myriad of social applications such as elderly care and education, it is important to find a reliable and robust interaction control strategy. This demand has brought about a major challenge for robot researchers and engineers.

In the literature of interaction control, two major frameworks are widely recognized, which are position/force control [5], [8], [12] and impedance control [7]. Compared to the former one, impedance control is more acceptable as it does not require the full decomposition of force and tracking trajectory. Besides, impedance control is more robust compared with position/force control. A passive impedance control will guarantee the stability if the environment is passive as well [3].

The performance of impedance control relies heavily on the proper selection of the targeted impedance. In the earlier research works, a desirable constant impedance is usually preferred and the researchers focused on how to deal with the uncertainties of the robot's dynamic model. These works include adaptive impedance control and learning impedance control as in [13], [2], [17], [11].

However, in many application scenarios, when the environment is totally unknown, passive impedance control may be too conservative if a high performance is required [1], [6]. To deal with this problem, impedance learning and optimization has been introduced. Learning and optimization is important in impedance control as the control objective includes both the trajectory tracking and force regulation, so an optimal or sub-optimal solution can be generated which is usually a trade-off of the two objectives. There have been a number of research studies in this area. In [9], the well-known linear quadratic regulator (LQR) optimal control is adopted for the proper selection of the impedance parameters. The environment model is assumed to be known and the optimal gain is calculated by solving the well-defined Riccati equation. However, the environment model is usually unknown, so the proposed method is simply not feasible for online implementation in practice.

To tackle this problem, adaptive dynamic programming (ADP) or actor-critic learning is proposed in [18], [19]. ADP mimics the way that biological system interacts with the environment. In the scheme of ADP, the system is considered as an agent which modifies its action according to the environment stimuli. The action is strengthened (positive reinforcement) or weakened (negative reinforcement) according to the evaluation of a critics. By using ADP, an optimal control policy can be generated with partial or none information of the system.

In this paper, we focus on the continuous critic learning for robot interacting with unknown environments. The proposed method is based on the research work in continuous ADP [10], where the optimal control solution is obtained subject to unknown system dynamics.

While the critic learning in [10] is only for the state regulation, it is further modified to handle the trajectory tracking. The developed impedance adaptation will result in the desired impedance parameters that are able to guarantee the optimal interaction, subject to unknown environments. Based on the above discussion, we highlight the contributions of this paper as follows:

(i) The unknown environment is considered in the analysis of the interaction control problem, which is defined as a linear system with unknown system dynamics.

(ii) The optimal control problem is modified such that the tracking problem is achieved using a regulation method and the desired impedance model can be obtained.

(iii) Continuous critic learning is adopted such that optimal impedance parameters in the sense of trajectory tracking and force regulation of robots are obtained subject to unknown environments.

The rest of the paper is organized as follows. In Section 2, the dynamics of environment are described, and impedance control and the objective of this paper are discussed. In Section 3, impedance adaptation based on continuous critic learning is developed for the described environment model, such that the optimal interaction is achieved subject to unknown environments. In Section 4, the validity of the proposed method is verified through simulation studies. Section 5 concludes this paper.

2. PROBLEM FORMULATION

2.1 Environment Modeling

In this paper, a damping-stiffness environment model (including human limb [16]) is considered, the dynamics of which can be described as follows

$$f = K_e x + B_e \dot{x} \tag{1}$$

where K_e and B_e are unknown damping and stiffness matrices; x and $\dot{x}$ are the robot arm's position and velocity; and f is the interaction force.

Remark 1: K_e and B_e are assumed to be unknown matrices in this paper. This assumption makes the problem studied in this paper more complicated compared with previous study in [14].

2.2 Impedance Control

Impedance control is first introduced in [7] to achieve certain desirable impedance and impose a desirable dynamic behavior to the interaction between the robot and environment. To apply the impedance control, we need to find a desired impedance model in the cartesian space as follows

$$f = g(x_d, x_v) \tag{2}$$

where x_d is the desired trajectory, x_v is the virtual desired trajectory in the Cartesian space, and $g(\cdot)$ is a target impedance function to be determined. Then, the virtual desired trajectory in the joint space $q_d = \int_0^t J^{-1}(q)\dot{x}_v(q)dq$ according to the interaction force f and the impedance model (2).

Remark 2: Eq. (2) is a general impedance model which defines the relationship between interaction force and position. In some specific applications, a simplified stiffness impedance model $G_{d1}x_v - G_{d2}x_d = -f$ can be adopted, where G_{d1} and G_{d2} are desired stiffness matrices.

2.3 Preliminary: Continuous Critic Learning

The continuous critic learning proposed in [10] is briefly introduced in the following, the results of which will be used for model-free impedance adaptation.

Consider the following continuous linear system

$$\begin{aligned} \dot{\xi} &= A\xi + Bu \\ y &= C\xi \end{aligned} \tag{3}$$

where ξ is the system state, y is the output, u is the system input, and A, B and C are unknown system matrices, subject to the following infinite-horizon optimal control problem

$$J = \int_0^\infty (\xi^T S\xi + u^T Ru)dt \tag{4}$$

where $S \geq 0$ and $R > 0$ are the weights of the state and input. For this system, it is well-known that the unique optimal control policy determined by the Bellman's optimal principle is given by

$$u = -K_{op}\xi \tag{5}$$

with $K_{op} = R^{-1}B^T P^*$, where the matrix P^* is obtained by solving the algebraic Ricatti equation (ARE)

$$A^T P^* + P^* A + P^* BR^{-1}B^T P^* + S = 0 \tag{6}$$

Continuous critic learning method is discussed in [10] to solve the Ricatti equation subject to unknown system parameters A, B, C. The procedure of the learning method is briefly introduced as follows.

Consider additional input dynamics

$$\varepsilon\dot{u} = v, u(0) = u_0 \tag{7}$$

which is perturbed by u and $\varepsilon > 0$ is a small constant.

Based on the system dynamics (3) and the additional input dynamics (7), the augmented linear system equation can be rewritten as

$$\dot{z} = Fz + Gv, z(0) = z_0 \tag{8}$$

where $F = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}, G = \begin{bmatrix} 0 & I_m/\varepsilon \end{bmatrix}^T$, $z = [\xi^T \ u^T]^T$ and I_m is the identity matrix.

The quadratic Q function, denoted by $Q_I(\xi(t), u(t))$ for the augmented system is given as

$$Q_I(\xi(t), u(t)) = \int_t^\infty (r(\xi, u) + v^T Rv)\, d\tau \tag{9}$$

where $r(\xi, u) = \xi^T S\xi + u^T Ru$.

There exists a unique solution $Q_I^*(\xi(t), u(t))$ which has a quadratic form as follows:

$$Q_I^*(\xi(t), u(t)) = [\xi^T \ u^T] \begin{bmatrix} H_{11}^\varepsilon & \varepsilon H_{12}^\varepsilon \\ * & \varepsilon H_{22}^\varepsilon \end{bmatrix} \begin{bmatrix} \xi \\ u \end{bmatrix} \tag{10}$$

where $H^\varepsilon = \begin{bmatrix} H_{11}^\varepsilon & \varepsilon H_{12}^\varepsilon \\ * & \varepsilon H_{22}^\varepsilon \end{bmatrix} \geq 0$ is the solution of the following Riccati equation

$$\begin{aligned} & F^T H^\varepsilon + H^\varepsilon + \Sigma \\ = & \begin{bmatrix} H_{12}^\varepsilon R^{-1}(H_{12}^\varepsilon)^T & H_{12}^\varepsilon R^{-1} H_{22}^\varepsilon \\ * & H_{22}^\varepsilon R^{-1} H_{22}^\varepsilon \end{bmatrix} \end{aligned} \tag{11}$$

where $\Sigma :=$diag$\{S, R\}$.

The approximated optimal control policy for (3) can be thus generated as

$$U(s) = -\Lambda(s)(H_{12}^\varepsilon)^T \xi(s) \tag{12}$$

where $\Lambda(s) = (\varepsilon sR + H_{22}^\varepsilon)^{-1}$ is the low-pass filter with the laplace variable s, and $U(s)$ and $\xi(s)$ are the laplace transforms of $u(t)$ and $\xi(t)$.

As mentioned in the Introduction, we aim to obtain optimal impedance parameters without the assumption that the environment model is given. This is the motivation to develop optimal impedance adaptation in this paper.

3. MODEL-FREE OPTIMAL IMPEDANCE ADAPTATION

3.1 Computational Neural Network Realization

The key problem of the continuous critic learning is to find the solution of the following temporal difference equation

$$\begin{aligned} Q_I^*(\xi(t), u(t)) = & \int_t^{t+T} (r(\xi, u) + v^T Rv)d\tau + \\ & Q_I^*(\xi(t+T), u(t+T)) \end{aligned} \tag{13}$$

In traditional ADP, this problem is usually solved in a dual actor-critic training structure and the temporal differential function is solved recursively. One major drawback of this approach is that the training process usually involves two approximators which make the structure of the training network very complicated.

Neural network has been acknowledged to have the excellent ability for universal approximation to any continuous model. This property may offer us a safer way to solve the problem. In the following, computational neural network will be adopted to solve the temporal differential equation.

The existing cost-function $Q_I^*(\xi(t), u(t))$ can be parameterized in the following forms

$$\begin{aligned} & Q_I^*(\xi(t), u(t)) \\ = \; & z(t)^T H^\varepsilon z(t) \\ = \; & (z(t)^T \otimes z(t))\text{vec}(H^\varepsilon) \\ = \; & (\text{vec}(H^\varepsilon))^T (z(t) \otimes z(t)) \end{aligned} \tag{14}$$

Similarly, the cost function from $t+T$ to ∞ can be derived as

$$\begin{aligned} & Q_I^*(\xi(t+T), u(t+T)) \\ = \; & z(t+T)^T H^\varepsilon z(t+T) \\ = \; & (z(t+T)^T \otimes z(t+T)^T)\text{vec}(H^\varepsilon) \\ = \; & (\text{vec}(H^\varepsilon))^T (z(t+T) \otimes z(t+T)) \end{aligned} \tag{15}$$

The above equation is important as it allows one to optimize over only one control vector at a time by working backward in time. In order to get the estimated value of the cost function, a conventional feed-forward network with one hidden layer and a linear output unit are constructed as follows

$$\hat{Q}_I(\phi, \omega) \;=\; \sum_{i=1}^{n_h} w_i^o g_i(\phi) + b^o \tag{16}$$

where "h" and "o" stand for "hidden" and "output" respectively; ϕ is the input of the network; $i = 1, 2, ..., n_h$ and n_h is the network number of the hidden layer; $g_i = tan(\sum_{j=1}^{n_l} w_{i,j}^h \phi(j) + b_i^h)$ denotes the hidden node output function where $j = 1, 2, ..., n_l$ and n_l is the network number of the input layer and w_i^o, $w_{i,j}^h$, b^o and b_i^h together form the network weights w.

It has shown that given the activation function g_i satisfying certain conditions, there exists a sequence of neural network function which approximates any given continuous target function. Recalling (14) and (15), if the temporal difference (13) is used for the training of neural network with $\phi = z(t) \otimes z(t)$, then we are able to approximate the existing quadratic cost function, i,e., $\hat{Q}_I(\phi, \omega) \to Q_I^*(\xi(t), u(t))$ as $t \to \infty$.

For the neural network given in (16), we can derive that $\theta = \text{vec}(H^\varepsilon)$ and the parameter θ can be computed by taking the partial derivative with respect to the corresponding input ϕ as below

$$\begin{aligned} \theta_j \;&=\; \frac{\partial \hat{Q}_I(\phi, \omega)}{\partial \phi_j} \\ &=\; \sum_{i=1}^{n_h} w_i^o w_{i,j}^h (1 + \tan^2(w_{i,j}^h \phi_j + b_i^h)) \end{aligned} \tag{17}$$

The vector ω of the NN weights in (16) can be obtained by error minimization between the target function using a back propagation algorithm.

Remark 3: Compared to traditional methods, the neural leaning method has the strength of fast training and only a very small data set is sufficient. In real-time environment where the system is shifting, this would prove to be a desirable property.

3.2 Optimal Impedance Adaptation

In this section, impedance adaptation will be further discussed. We will first show how to transform a tracking problem into a regulation problem and how to integrate the continuous critic learning discussed in Sections 3.1 and 3.2 into the impedance control in Section 2.2. Under this scheme, the targeted impedance is adapted during the manipulation process which achieves an approximated optimal performance.

For the damping-stiffness environment described in Section 2.1, the following cost function can be developed

$$J = \int_0^\infty ((x - x_d)^T S_1 (x - x_d) + f^T R_1 f) dt \tag{18}$$

where x_d is the desired continuous trajectory; S_1 is the weight of the trajectory tracking error, and R_1 is the weight of the interaction force. Besides, $S_1 = S_1^T \geq 0$ and $R_1 = R_1^T \geq 0$.

As shown in (18), the optimal problem is in fact a tracking problem, which is concerned with making the system output follow or track a desired trajectory. How-

ever, the traditional optimal problem is usually a regulation problem which can be regarded as a special case where the desired trajectory is zero state. Therefore, some manipulations are needed to make the problem identical. In particular, we consider

$$\eta = [x \quad p]^T \tag{19}$$

where p is the state of the following system

$$\begin{cases} \dot{p} = Up \\ x_d = Vp \end{cases} \tag{20}$$

where U and V are two known matrices and (U, V) is observable.

Considering the system in Section 2.1, if the interaction force f is considered as the system input u to the environment, the system dynamics described in (3) can be applied where $\xi = x, A = -{B_e}^{-1}K_e, B = -{B_e}^{-1}$ and $C = 1$.

As the formulated state p is observable, the augmented state and state matrix can be defined as follows

$$\begin{aligned} \hat{A} &= \begin{bmatrix} A & 0 \\ 0 & U \end{bmatrix}, \hat{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \\ \hat{S} &= \begin{bmatrix} C^T S_1 C & -C^T S_1 V \\ -V^T S_1 C & V^T S_1 V \end{bmatrix}, \hat{R} = R_1 \end{aligned} \tag{21}$$

The system thus can be re-written in the following state-space form

$$\dot{\eta} = \hat{A}\eta + \hat{B}f \tag{22}$$

Then the infinity cost is obtained as

$$J = \int_0^\infty (\eta^T \hat{S} \eta + f^T \hat{R} f) dt \tag{23}$$

The system now has the same form as discussed in Section 2.3, so the continuous critic learning method can be adopted. It is trivial to show that the following optimal control policy can be obtained as

$$f = -K^* \eta \tag{24}$$

where K^* is calculated using the methods described in Sections 2.3 and 3.1. The exact impedance function which guarantees the optimal interaction is thus obtained. Recalling the impedance control as discussed in Section 2.2, the desired targeted impedance is achieved according to the measured f and given p, and the inner loop is to guarantee the trajectory tracking in the joint space.

4. SIMULATION STUDY

To verify the proposed impedance adaptation, in this section, a robot manipulator with two-degrees-of-freedom is considered. The damping-stiffness environment model described in Section 2 is adopted to describe a typical environment. The simulation is performed using the robotics toolbox [4].

The parameters of the robot arm are given as $m_1 = m_2 = 2.0$kg, $l_1 = l_2 = 0.2$m, $i_1 = i_2 = 0.027$kgm^2, $l_{c1} = l_{c2} = 0.1$m, where m_j, l_j, i_j, l_{cj}, $j = 1, 2$, represent the mass, the length, the inertia about the z-axis that comes out of the page passing through the center of mass, and the distance from the previous joint to the center of mass of the current link, respectively.

It is assumed that the environment force is only exerted to the robot arm along the x axis and the y axis is interaction free. Adaptive control discussed in [15] is adopted to guarantee the inner loop control performance.

Corresponding to Section 2.1, the following environment is considered: $0.01\dot{x} + 4.1(x - 0.2) = -f$. Note that 0.2 is the initial position of the robot arm. As the environment is known in the simulation, the exact optimal solution and desired impedance model can be obtained by solving the Riccati equation which is referred to as "LQR", and compared with the the proposed method in this paper which is referred to as "Proposed".

It is necessary to emphasize that the environment dynamics are only available in the simulation and they are usually unknown or need to be estimated in real applications. This is the motivation of this paper, which has already been discussed in the Introduction.

The weights in (18) are given by $S_1 = 1$ and $R_1 = 1$. The desired impedance model is $f = -0.1202x + 0.2364x_d$ based on known A and B. The simulation results are shown in Figs. 1, 2 and 3. Using the impedance adaptation, the desired impedance model is obtained as

$f = -0.1274x + 0.2369x_d$. It is found that the obtained impedance model is very near to but not exactly the same as the desired one under LQR. This may be caused by the adaptation process in the inner-loop, as the perfect tracking cannot be fully guaranteed. Therefore, the proposed method only realizes "sub-optimal" impedance control if the "perfect" tracking in the inner-loop cannot be guaranteed.

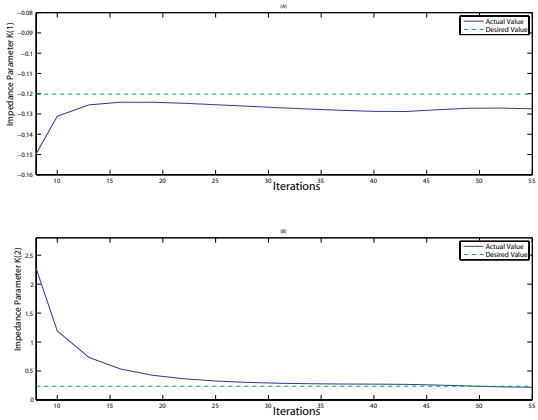

Fig. 1 Desired and adapted impedance parameters, $S_1 = 1$ and $R_1 = 1$

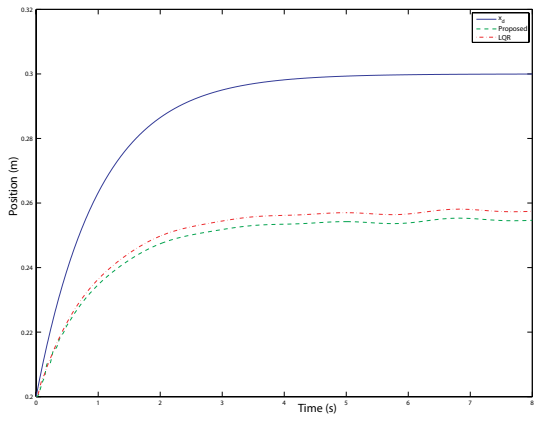

Fig. 2 Desired and actual trajectory, $S_1 = 1$ and $R_1 = 1$

The interaction force is shown in Fig. 3 to further illustrate the validity of the proposed method.

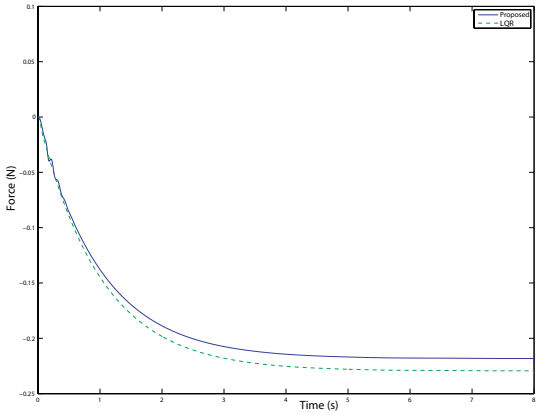

Fig. 3 Interaction force, $S_1 = 1$ and $R_1 = 1$

To further illustrate the effectiveness of the proposed method, another cost function is chosen in the second case. The weights in (23) are given by $S_1 = 3$ and $R_1 = 1$. Compared to that in the first case, the weight of the tracking error is larger, so it is expected that the tracking error becomes smaller and interaction force becomes larger. Similarly, the desired impedance model is obtained as $f = -0.1552x + 0.3200x_d$ based on known A and B. The simulation results in this case are given in Figs. 4, 5 and 6, and the impedance model obtained with the proposed method is $f = -0.1625x + 0.2872x_d$. The simulation results further confirm the validity of the proposed method.

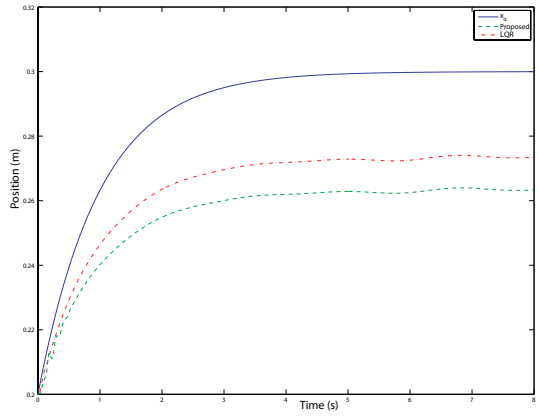

Fig. 4 Desired and actual trajectory, $S_1 = 3$ and $R_1 = 1$

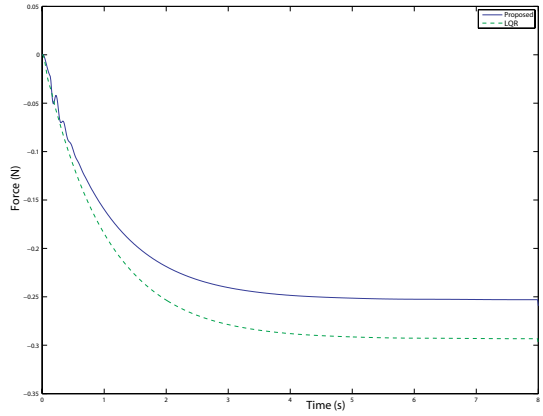

Fig. 5 Interaction force, $S_1 = 3$ and $R_1 = 1$

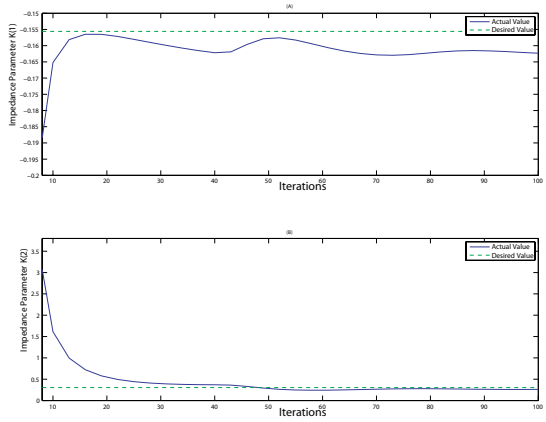

Fig. 6 Desired and adapted impedance parameters, $S_1 = 3$ and $R_1 = 1$

5. CONCLUSION

In this paper, continuous critic learning has been employed for optimal impedance adaptation subject to unknown environments. The unknown environment is modeled as a linear system with unknown system parameters and a certain cost function which combines trajectory tracking and force regulation has been adopted. The optimal impedance model has been obtained using the computational neural network. The validity of the proposed methods has been verified through simulation studies.

REFERENCES

[1] S. P. Buerger and N. Hogan. Complementary stability and loop shaping for improved human-robot interaction. *IEEE Transactions on Robotics*, 23(2):232–244, 2007.

[2] C. C. Cheah and D. Wang. Learning impedance control for robotic manipulators. *IEEE Transactions on Robotics and Automation*, 14(3):452–465, 1998.

[3] J. E. Colgate and N. Hogan. Robust control of dynamically interacting systems. *International Journal of Control*, 48(1):65–88, 1988.

[4] P. I. Corke. A robotics toolbox for MATLAB. *IEEE Robotics and Automation Magazine*, 3(1):24–32, March 1996.

[5] J. J. Craig and M. H. Raibert. A systematic method of hybrid position/force control of a manipulator. *Computer Software and Applications Conference, IEEE Computer Society*, pages 446–451, 1979.

[6] V. Duchaine and C.M. Gosselin. Investigation of human-robot interaction stability using Lyapunov theory. In *IEEE International Conference on Robotics and Automation, Piscataway, NJ, United States*, pages 2189–2194, 2008.

[7] N. Hogan. Impedance control: an approach to manipulation-part I: Theory; part II: Implementation; part III: Applications. *Transaction ASME J. Dynamic Systems, Measurement and Control*, 107(1):1–24, 1985.

[8] L. Huang, S. S. Ge, and T. H. Lee. Position/force control of uncertain constrained flexible joint robots. *Mechatronics*, 16(2):111–120, 2006.

[9] R. Johansson and M. W. Spong. Quadratic optimization of impedance control. *Proceedings of IEEE International Conference of Robotics and Automation*, 1:616–621, 1994.

[10] Jae Young Lee, Jin Bae Park, and Yoon Ho Choi. Integral q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems. *Automatica*, 2012.

[11] Y. Li, S. S. Ge, and C. Yang. Learning impedance control for physical robot-environment interaction. *International Journal of Control*, 85(2):182–193,

2012.

[12] Z. J. Li, S. S. Ge, and A. Ming. Adaptive robust motion/force control of holonomi-constrained nonholonomic mobile manipulators. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, 37(3):607–616, 2007.

[13] W.-S. Lu and Q.-H. Meng. Impedance control with adaptation for robotic manipulations. *IEEE Transactions on Robotics and Automation*, 7(3):408–415, 1991.

[14] M. Matinfar and K. Hashtrudi-Zaad. Optimization-based robot compliance control: Geometric and linear quadratic approaches. *The International Journal of Robotics Research*, 24(8):645–656, 2005.

[15] J. J. E. Slotine and W. Li. On the adaptive control of robotic manipulators. *The International Journal of Robotics Research*, 6(3), 1987.

[16] T. Tsumugiwa, R. Yokogawa, and K. Hara. Variable impedance control based on estimation of human arm stiffness for human-robot cooperative calligraphic task. *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 644–650, 2002.

[17] D. Wang and C. C. Cheah. An iterative learning-control scheme for impedance control of robotic manipulators. *The International Journal of Robotics Research*, 17(10):1091–1099, 1998.

[18] P. J. Werbos. Using ADP to understand and replicate brain intelligence: The next level design. *Proceedings of the 2007 IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning*, pages 209–216, 2007.

[19] P. J. Werbos. Intelligence in the brain: A theory of how it works and how to build it. *Neural Networks*, 22(3):200–212, 2009.